Q: Why does EM work?

Let $f(y,z;\theta)$ be the density for the complete data (y,z)

Recall $f(y;\theta) = \int f(y,z;\theta)\,dz$ is the marginal density of Y.

log-likel: $\ell(\theta) = \ln f(y;\theta)$

Let θ^ν be the current value. Calculate difference $\ell(\theta) - \ell(\theta^\nu)$

$$\ell(\theta) - \ell(\theta^\nu) = \ln f(y;\theta) - \ln f(y;\theta^\nu)$$

$$= \ln \int f(y,z;\theta)\,dz - \ln f(y;\theta^\nu)$$

$$= \ln \int f(y,z;\theta) \cdot \frac{f(z|y;\theta^\nu)}{f_{Z|Y}(z;y,\theta^\nu)}\,dz - \ln f(y;\theta^\nu)$$

multiply/divide by conditional density $Z|Y$ when $\theta = \theta^\nu$

$$= \ln \int \frac{f(y,z;\theta)}{f(z|y;\theta^\nu)}\, f(z|y;\theta^\nu)\,dz - \ln f(y;\theta^\nu)$$

$$= \ln E_{Z|Y;\theta^\nu}\left[\frac{f(Y,Z;\theta)}{f(Z|Y;\theta^\nu)}\right] - \ln f(y;\theta^\nu)$$

[ln is "concave" fn $\Rightarrow$ Jensen's ineq $\ln EX \geq E \ln X$]

$$\geq E_{Z|Y;\theta^\nu}\left(\ln \frac{f(Y,Z;\theta)}{f(Z|Y;\theta^\nu)} \cdot \frac{1}{f(Y;\theta^\nu)}\right)$$

$$= E_{Z|Y;\theta^\nu}\left[\ln \frac{f(Y,Z;\theta)}{f(Y,Z;\theta^\nu)}\right]$$

In short

$$\ell(\theta) - \ell(\theta^{\nu}) \geq \underbrace{E_{Z|Y;\theta^{\nu}}\left[\ln \frac{f(Y,Z;\theta)}{f(Y,Z;\theta^{\nu})}\right]}_{\text{when } \theta=\theta^{\nu} \Rightarrow \text{RHS}=0.} \cancel{= Q(\theta;\theta^{\nu},y)}$$

Max RHS is equivalent to maximizing $Q(\theta;\theta^{\nu},y)$

Thus if we choose $\theta = \theta^{\nu+1}$ that maximizes the RHS it is fair to say that

$$E_{Z|Y;\theta^{\nu}}\left[\ln \frac{f(Y,Z;\theta^{\nu+1})}{f(Y,Z;\theta^{\nu})}\right] \geq 0$$

$$\Rightarrow \ell(\theta^{\nu+1}) - \ell(\theta^{\nu}) \geq 0$$

Note that maximizing the RHS wrt θ really means maximizing

$$E_{Z|Y;\theta^{\nu}}\left[\ln f(Y,Z;\theta)\right] \text{ wrt } \theta.$$

This is exactly the "Q"-function.

Choosing the updates for θ in this way guarantees that likel increases.

qed.

Last remarks

θ^{ν} may be a stationary point which need not be local max

In some cases EM converges to a local minimo or a saddlepoint

However if the likel fn is unimodal and satisfies some differentiability conditions, then EM converges to the unique MLE.